ciência^enatura

A new method to classify breast cancer tumors and their fractionation

Omid Rahmani-Seryasat[1], Javad Haddadnia[2,*],Hossein Ghayoumi-Zadeh[1]

[1] Department of Electrical and Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran

[2] Associate Professor, Electrical and Computer Engineering Department, Hakim Sabzevari University, Sabzevar, Iran

Abstract

In this paper, suspicious breast tumors were classified by using the neural network and the growth area method has been used for a fractionation of the benign or malignant areas of the normal tissue. Features extracted from input tissues are including statistical features and characteristics of spatial dependence. The advantage of this method is using of phase adaptive threshold based on entropy which leads to more accurate extraction of tumors and also corresponded with the nature of mammogram images. As a result, this method mimics of the human eye operation to detect abnormal masses. Database used in this paper is the MIAS mammogram database including 238 normal, benign and malignant mammograms. The accuracy obtained with 38 features is equal to 86.66% for detecting abnormal masses and 38.05 % for normal masses.

1 Introduction

Computer analysis methods of mammographic images which are based on image processing techniques have been used in various fields, including the improvement of mammographic features, detecting and locating suspicious areas and classification of the micro-calcification clusters. Many of the methods in micro-calcification computer analysis, fractions mammogram to binary areas at first which each area fractioned image showing a micro-calcification. Then the feature analyze would be done on these binary images. Calcification clusters shapes and sizes can be very possible to be associated with cancer complications. Also gray levels values which represents local structure of clusters, have an important role in micro-calcification actual interpretation.

Researchers have different methods for extracting cancer masses. The LI fraction the suspicious areas by fractionation adaptive. Petrik suggested fractions based on improving contrast according to the density and Zheng recommended high-pass filter for the mass fractions. In this paper, a back propagation neural network with normal and abnormal input samples are taught. At new mammogram, suspicious areas are extracted by using growth areas and neural network determine the mass type by using features of the suspected areas.

In section 2, the method used for preprocessing input mammogram, feature extracted, the neural network used, and suspicious mass fractionated method have been described. In Section 3 the test results has been provided and in section 4 they have been concluded.

The proposed method

Automatic detection system used in this article has been shown in figure 1. Input mammogram in the first step will be improved by using the histogram integration methods and morphology operators. In the next step, Statistical features and pixels spatial correlation are extracted. Then

classifier with these features will be trained and finally the test would be done on the new mass.

In the test step, the areas placed in abnormal classes by the neural network, as the core are given to the growth suspected areas method and then by applying competitive phase adaptive based on entropy noise removal and extracting masses would be done.
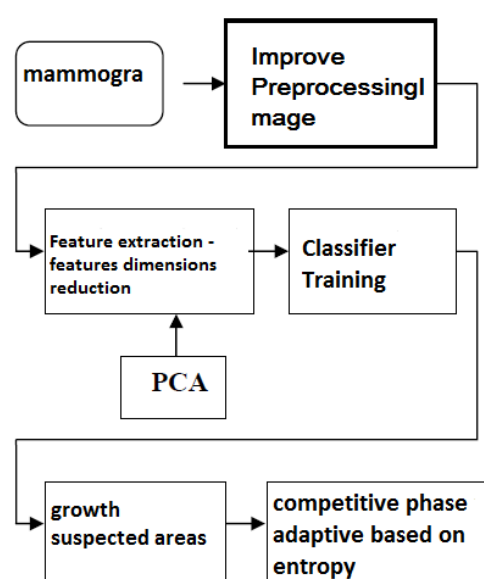

Figure 1: diagram used in automatic detection system of cancer masses

1.2 Improving original image

2.2 A) integration of the histogram

If we consider the histogram of a mammogram as a probability distribution, according to information theory, Mammograms produced uniform histogram, have more information itself. If the histogram were uniform as much as possible, to the greatest amount of entropy would be reached and mammogram image would be improved. To uniform histogram to the fractionated areas independently the Histogram adaptive integration method can be used. However some normal tissues and noises would be improved [2].

3.2 morphology Operators

Morphology methods and defining the border-operators can be used to make improvement. This process is performed by making improvements in the features which are similar to point and through the morphology process Top-hat. This operator is used to scan objects with different light background in the images with a non-uniform lighting background. Formula used for this function is as follows:

$B = A − (Ao\ SE)$

Which A is the input image, B is the preprocessing image, SE is the structure element, o is the opening operator and – is the two images difference operator. After the difference image operation was performed, the Top of the original image A represent the area we are looking for.

4.2 Feature Extraction

A mammogram usually involves a lot of diverse information which represents the different tissues, blood vessels, skin, breast, the edge of the breast, mammography film, and features related to the imaging device [7]. To obtain a good detection system which can categorize normal or abnormal mammograms areas, we have to give all the mammograms information to the recognition system so the system is able to recognize normal and abnormal tissues correctly. However, the use of diverse information Leads to obtain high-dimensional feature vectors and reduce identify accuracy and increases the computational load. Therefore by eliminating unrelated information, you can get reliable mammogram descriptors [4]. In this article, the statistical characteristics and the occurrence matrix features have been used to describe and classify different tissues.

5.2 a) statistical properties of tissue

According to the idea that the masses of the same kind the of their neighbors, have the same tissue features, the tissues statistical features can be used for Classification of normal, malignant and benign tissues. Features used in this article includes the mean, standard deviation, smoothness, third moment, Uniformity and Entropy [6.3].

In the figure 2 some examples of normal and abnormal tissues have been shown in the image mdb028.pgm. The histogram correspond to these

tissues is shown in Figure 3. It is evident that the abnormal tissue histogram of brighter gray levels has more value than normal.
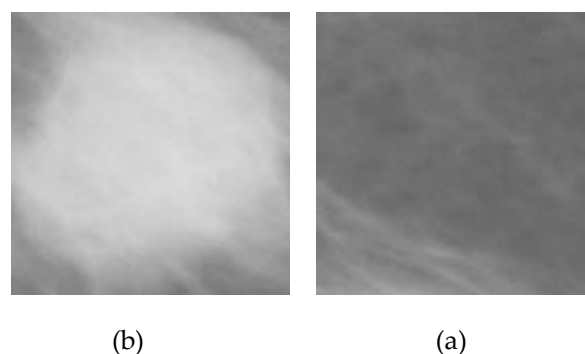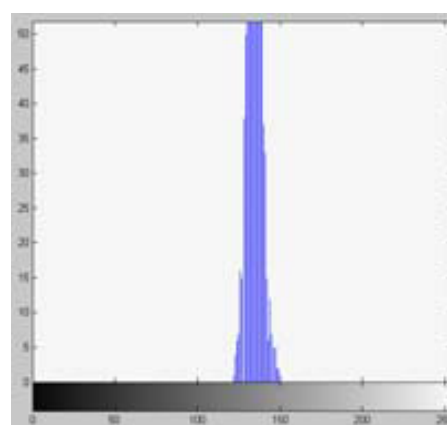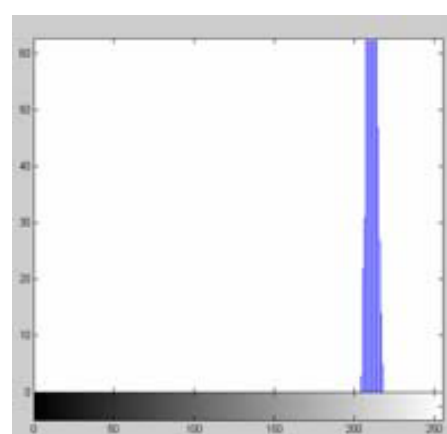

(b)   (a)

Figure 2: (a) is the normal tissue, (b) the abnormal tissue of mammogram mdb 028.pgm

(a)

(b)

Figure 3. Image 4 histogram (a) normal tissue histogram, (b) abnormal tissue histogram

6.2 Use of the occurrence matrix

The first level histogram of gray levels, gives the probability distribution of a gray area occurrence in the image. The second level histogram is

shown with H(yq, yr, d) which in some references, also known by the name of the occurrence matrix. This probability distribution matrix show the occurrence of a pair of gray levels which their position are determined to each other by the displacement vector d. In the other words represents the number of times that the gray levels couple $H(y, y, d)$ have been separated, with vector d. second level statistic features defined the overall structure of the image very well[1.4].

7-2 Category

8-2 A) Neural Networks

In the case that the Knowledge Expert System not expressed clearly and cannot be expressed in the form of self-rule regions, neural networks would be used widely. Artificial neural networks can be used at mammograms reviews. The neural network used here is the Back-Propagation Nueral Network.

9-2 Fuzzy clustering

Fuzzy clustering is the process of classifying a set of unlabeled patterns to some clusters so that similar patterns attributed to a cluster. The hard clustering techniques, compared each member of the data exactly to a cluster. Algorithm FCM and its derivatives are used successfully in many cases, such as pattern recognition, classification, data mining, and scan images. It is also used for data analysis and model. Usually the FCM consists of several administrative steps. In the first step, the initial cluster center selected randomly of input data. Then, on the next steps, after successive iterations of the algorithm, the result will converge to the actual data clusters. As a result, the selection of the initial set of cluster centers is very important for the algorithm FCM. By the way choosing a good set of initial cluster centers is difficult. [9.10.11]

In this article for fractions breast of background, algorithm FCM has been used with the euclidean distance.

10-2 Mass fractions based on adaptive neighborhood

First those who use the comparative neighborhood to improve the contrast of

mammogram images were Gordon and Rangayyan [4]. The suggested method included growth background. Morrow improved this method by a new definition of the background areas [7]. The algorithm improves the contrast comparative neighborhood acts as follows: Initial core, pixels is with a maximum amount of light. Other pixels in addition to have more high brightness compared to other pixels, must have a minimum distance of other cores. This condition Prevent of created core as aggregation in an area. The process to develop the areas of connected pixels and similar to core are be considered as the background. Similar to the pixel brightness difference means less than a certain limit. This value is considered in the implementation stage equal to 10. This growth continues until Tape width 3 pixels be obtained for the background. Background pixels are pixels that are not similar to the core. The new value of the cells pixel is obtained on the basis of the contrast between the foreground and background areas [3.5]

2.11 Threshold-based entropy

In the next step, first we determine the membership degree of each pixel to the class of the field, noise and edge by the phase competition method and eliminate the possible noises. Then in the windows 10×10 entropy change curve is plotted for different thresholds and at the peak, the threshold is performed. As a result, the threshold of as comparative is performed for areas with different lighting. So cancerous areas (True Positive) that have been diagnosed correctly would increases and the number of areas that go wrong in cancer group (False Positive) reduced.

3. Test results

To evaluate the proposed method of database MIAS, Mini Mammographic has been used. Mammogram are with $200\mu m$ resolution and dimensions 1024×1024. There are usually about 119 abnormal areas at this database consisting varies abnormal tissues. The smallest mass in this database with the radius of 3 pixels, and the largest has the radius of 197 pixel. The normal tissues have been selected from the normal breasts accidently [8].

The database MIAS specify each mammogram with a circle. The information includes an

estimate of the mass and radius of it. Since the mass of non-normal is not necessarily circular, the masses provided with this view can be a normal tissue. Therefore for training and test, the areas with the size 35 × 35 to the center abnormal mass, have been chosen because the mass center has the most importance.

1-3 Classification with statistical features

Of the 238 areas existing in the MIAS database, 119 areas have been used for training which there are 60 abnormal areas and 59 normal areas among them. 6 Statistical area features evaluated after normalization was given to the nervous network input layer. Total mass of the second and third layers of the nervous network were respectively6 and 3. The average correct diagnosis in the training step is equal to 96.6% (4 error in 119 samples). For the number of layers in the testing phase, it classified wrongly to 8 abnormal areas and 10 normal areas (table 1).

2.3 Classification by characteristics of the occurrence

For the characteristics of the place, d=1, -45 ° angles, 0°and 45° have been used. Neural network architecture for the previous section, for the angle 45° 11 normal areas and 10 abnormal areas, for the angle 0° 15 normal areas and 14 abnormal areas, for the angle -45 ° 12 normal areas and 11 abnormal areas were classified wrongly(table 2).

Table 1. The diagnostic accuracy using statistical features

Mass type	The percentage of correct diagnosis
abnormal	86.66%
normal	83.05%
Sum	84.85%

Table 2. Diagnostic accuracy for the spatial autocorrelation characteristics.

The angle of the occurrence matrix	normal	abnormal	Sum
45°	81.35%	83.33%	82.34%
0°	74.75%	76.66%	75.61%
− 45°	79.66%	81.66%	80.66%

Learning curve for the 38 features with the neural network is shown in Figure 4. After 32 APEC the training error has arrived zero.
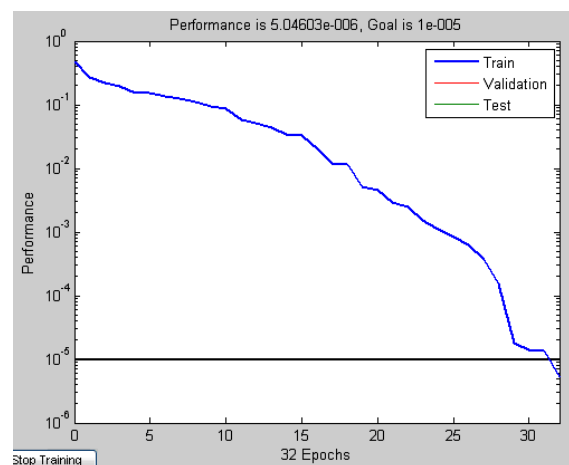

Figure 4: Error changes curve in neural network training for the 38 Features

Test results are shown in Table 3, taking into account all the features and PCA actions to reduce the dimensions of the features.

Table 3. The diagnostic accuracy to reduce the size of features to help PCA

The number of Features	Normal	abnormal	Sum
85.39	86.66	84.13	38
78.27	77.42	79.12	30
78.43	77.42	79.45	25
75.34	71.33	79.35	20
75.58	72.12	79.05	15
77.54	74.35	80.74	5
84.23	84.33	84.13	10

3.3 Mass extraction

For the images fractionated and obtain the exact location of the mass, the growth method based on the area has been used. The best answer has obtained for the total of 20 primordial nucleus and the connection equal to 10. Cells were selected in areas which mask the tissue fractions were diagnosed suspected tumors. In Figure 5, the result of fractions image using an algorithm the growth of the areas has been shown. The image 5(a) shows the mdb028.pgm. in the image 5(b), the results of breast fractionation of the background has been shown with the help of FCM algorithm. Breast fractionation of the background help to increase calculating speed.

The image 5(c) is the improved image by histogram integration and operator morphology Top-hat. The image 5(d) represents the algorithm applied to develop the areas of adaptive threshold based on the maximum entropy on the image.

4. Discussion

In the suggested method due to using the threshold adaptive fuzzy which works similar to the human visual system, diagnostic power of in areas with low contrast has been increased dramatically. An important advantage of the human eye toward intelligent systems, is the

micro calcification detection in dense areas. With the adaptive threshold, reduce the threshold value in the areas of clear (Which equates to identify clusters in dense areas as result of increase in TP). And increase it in the dark areas which leads to FP reduction. Actually we are do it in the opposite of the background lightening. The use of fuzzy methods to fractionation the chest area, matches the uncertain nature of breast tissue.

Also mask applied to fractionation tissue with large steps, then applying the more accurate it in suspected areas can help to increase the speed of proposed method.

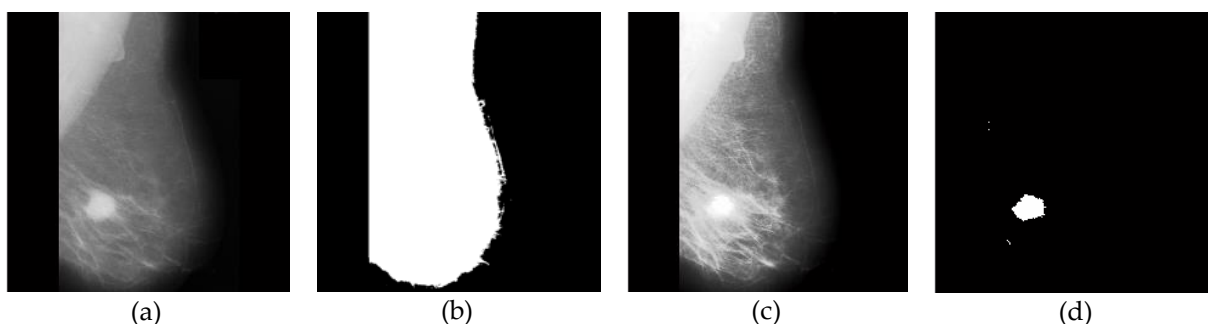| (a) | (b) | (c) | (d) |

Figure 5. Image fractionation results (a) mdb028.pgm. The original image b is the result of breast fractionation of background (c) improved image (d) fractioned mass with growth areas and the adaptive threshold.

The use of boundary features surrounding masses of torque and the edge of the histogram can be used to improve the proposed method. You can also use methods based on genetic algorithms to determine the best features.

5. Conclusion

In this paper, a system for classification, fractionation of cancerous mass was presented and the performance of different features extracted from the masses were studied. PCA method was used to reduce the size of the features and it is evident that the reduction in size will not compromise the accuracy of the detection system. Use angle ° 45 to achieve the characteristics of the place has obtained the best results and increase the accuracy. The growth areas technique has been used to fractionation images which its results shows that this method has the best adapted to the actual location of the masses.

References

Farshid Rafiee "Mammographic image analysis and classification of calcium deposits using multimedia features Violet" Artificial Intelligence Master Thesis, Department of Computer Engineering,. Amir Kabir Industrial University, 1999.

H.D.Cheng,Xiaopeng Cai,Xiaowei Chen,Liming Hu,Xueling Lou "Computer-aided detection and classfication of microcalcications in mammograms: a survey", Pattern Recognition, vol 36, pp. 2967-2991, 2003.

J.C. Fu, S.K. Lee, S.T.C. Wong, J.Y. Yeh, A.H. Wang, H.K. Wu, "Image segmentation features selection and pattern classification for mammographic microcalcifications ", Computerized Medical Imaging and Graphis, vol. 29, pp. 419-429, 2005.

I.Christoyianni, A. Koutras, E. Dermatas, G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms" –

Computerized Medical Imaging and Graphics, vol.26, pp. 309-319, 2002.

H.D. Cheng, X.J.Shi, R.Min, L.M. Hu, X.P. Cai, H.N. Du, " Approaches for automated detection and classification of masses in mammograms ", Pattern Recognition, vol. 39, pp. 436-486, 2006.

Sheshadri, A. Kandaswamy "Experimental investigation on breast tissue classification based onstatistical feature extraction of mammograms", Computerized Medical Imagings and Graphics, vol.31, pp. 46-48, 2007.

M.R.M. Samulski, "Classification of Breast Lesions in Digital Mammograms", Master Thesis, Radboud University Nijmegen,2006.

http://peipa.essex.ac.uk/info/mias.html

Miin-ShenYang, Kuo-LungWu, 'Unsupervised possibilistic clustering', Pattern Recognition, vol. 39 pp.